UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
Department of Electrical and Computer Engineering

ECE 544NA Pattern Recognition
Fall 2014

**EXAM 2**

Thursday, November 6, 2014

- This is a CLOSED BOOK exam. You may use one page, both sides, of notes.
- There are a total of 100 points in the exam. Plan your work accordingly.
- You must SHOW YOUR WORK to get full credit.

| Problem | Score |
|---|---|
| 1 | |
| 2 | |
| 3 | |
| Total | |

**Name:** ______________________________________________

**Problem 1 (34 points)**

Amanda has just passed her ECE qualifying exam. To celebrate, she gets a little bit drunk. Fortunately, it is very easy for her to find her way home, because her house is situated at the spot in Urbana with the lowest altitude. She devises the following algorithm to get herself home. For times $k \in \{1, 2, \ldots\}$ as necessary until arrival at home, she repeats the following process:

(a) Choose a direction, uniformly at random. Lean in that direction.

(b) If the ground slopes downhill, take a step.

(c) If the ground slopes uphill, take a step with probability $P(z_{k-1}, z_k)$ dependent on the current altitude $z_{k-1}$ and proposed future altitude $z_k$, where

$$P(z_{k-1}, z_k) = e^{-\alpha(z_k - z_{k-1})/T_k} \tag{1}$$

The altitudes $z_{k-1}$ and $z_k$ are measured in meters above sea level; as you may know, altitudes in Urbana vary in the range $200 \leq z \leq 250$ meters (Amanda's house is at $z^* = 200$ meters). The variable $T_k$ represents the tempo, in beats per minute, of the music on her mp3 player; since her battery is running out, this is a strictly non-increasing function of time, $T_{k-1} \geq T_k$. The constant $\alpha = 10$ bpm/meter.

(a) Given the information that you have, what are the smallest possible values of $T_k$ that are guaranteed, with probability one, to eventually deliver Amanda to her house?

(b) Suppose that Amanda starts her journey in a bar at altitude $z_0 = 240m$. Is this information sufficient to change your answer to part (a)? Why or why not?

(c) Suppose that now Amanda has developed the ability to teleport. Therefore her algorithm for getting home is now, for $k \in \{1, 2, \ldots\}$,

(i) Choose a target location somewhere in Urbana, uniformly at random.

(ii) If that location is downhill, teleport there.

(iii) If that location is uphill, teleport there with probability $P(z_{k-1}, z_k)$ given by Eq. 1.

Under this algorithm, what are the smallest possible values of $T_k$ that guarantee, with probability one, that Amanda will eventually find her way home?

(d) Suppose that we fix $T_k$ at a constant slow tempo, $T_k = 2$bpm, for all time. Prove that, under this circumstance, Amanda is guaranteed to eventually reach home with probability one (although her algorithm will not guarantee that she stays there once she has arrived).

**Problem 2 (33 points)**

Consider an RBM with visible variables $v \in \{0,1\}^p$, and with hidden variables $h \in \Re^q$. Suppose that the joint probability $p(h,v)$ is given by

$$p(h,v) = \frac{1}{Z} e^{-h^T W v - \frac{1}{2} h^T R h} \tag{2}$$

where $R$ is positive definite, and $W = [w_1, \ldots, w_p]$ for column vectors $w_k = [w_{1k}, \ldots, w_{pk}]^T$.

(a) Suppose that, for some odd reason, you have a training database in which both $h_i$ and $v_i$ are specified. Define

$$\mathcal{L} = \sum_{i=1}^{n} \ln p(h_i, v_i)$$

Compute $\partial \mathcal{L} / \partial w_{jk}$ as a function of $h_1, \ldots, h_n$, $v_1, \ldots, v_n$, $W$, and $R$. Your answer may include unevaluated explicit integrals or sums, but should not include any unevaluated expectations.

(b) What is $E[h|v]$? Remember that $h \in \Re^q$. Specify your answer in terms of $v$, $W$, and $R$. Your answer should *not* include any unevaluated integrals, sums, or expectations. You may find it useful to know that every positive definite matrix, $A$, has a square root ($A = A^{1/2}A^{1/2}$), and an inverse ($AA^{-1} = I$), and that both $A^{1/2}$ and $A^{-1}$ are also positive definite.

(c) Now imagine the converse: you are given $h$. Remember that $v \in \{0,1\}^p$. Please compute the conditional probability that the $j^{\text{th}}$ component of $v$ is "turned on," that is, compute $\Pr\{v_j = 1|h\}$. Your answer should *not* include any unevaluated integrals, sums, or expectations.

**Problem 3 (33 points)**

This problem considers a few different function classes that map from $\mathcal{X} = \Re^p$ to $\mathcal{Y} = \{0, 1\}$.

(a) Consider the function class

$$\mathcal{H}_a = \{h : h(x) = [x == \mu]\,, \quad \mu \in \Re^p\}$$

where $[\cdot]$ is the unit indicator function. What is the VC dimension of class $\mathcal{H}_a$? Prove your answer: demonstrate that, with probability one, a training dataset $mathcalD = \{x_1, \ldots, x_n\}$ of size $n > d$ can be labeled in only $\mathcal{O}\left\{n^d\right\}$ different ways for VC dimension $d$.

(b) Consider the function class

$$\mathcal{H}_b = \left\{ h : h(x) = [\|x - \mu\| < b] \,, \quad \mu \in \Re^p, \quad b \in \Re^+ \right\}$$

where $[\cdot]$ is the unit indicator function, and $\|x\|$ is the Euclidean norm. What is the VC dimension of class $\mathcal{H}_b$? Prove your answer: demonstrate that a training dataset $mathcal D = \{x_1, \ldots, x_n\}$ of size $n > d$ can be labeled in only $\mathcal{O}\left\{n^d\right\}$ different ways.

(c) Consider the function class

$$\mathcal{H}_c = \left\{ h : h(x) = \left[ \mu^T x \geq 1 \right], \quad \mu \in \Re^p \right\}$$

where $[\cdot]$ is the unit indicator function. What is the VC dimension of class $\mathcal{H}_c$? Prove your answer: demonstrate that a training dataset $mathcal D = \{x_1, \ldots, x_n\}$ of size $n > d$ can be labeled in only $\mathcal{O}\left\{ n^d \right\}$ different ways, and/or demonstrate that a dataset can be labeled in $2^n$ different ways if and only if $n < d$.